VEmotion: Using Driving Context for Indirect Emotion Prediction in Real-Time

David Bethge
Porsche AG, LMU Munich
Stuttgart, Germany

Thomas Kosch
TU Darmstadt
Darmstadt, Germany

Tobias Grosse-Puppendahl
Porsche AG
Stuttgart, Germany

Lewis L. Chuang
TU Dortmund
Dortmund, Germany

Mohamed Kari
Porsche AG
Stuttgart, Germany

Alexander Jagaciak
Porsche AG
Stuttgart, Germany

Albrecht Schmidt
LMU Munich
Munich, Germany

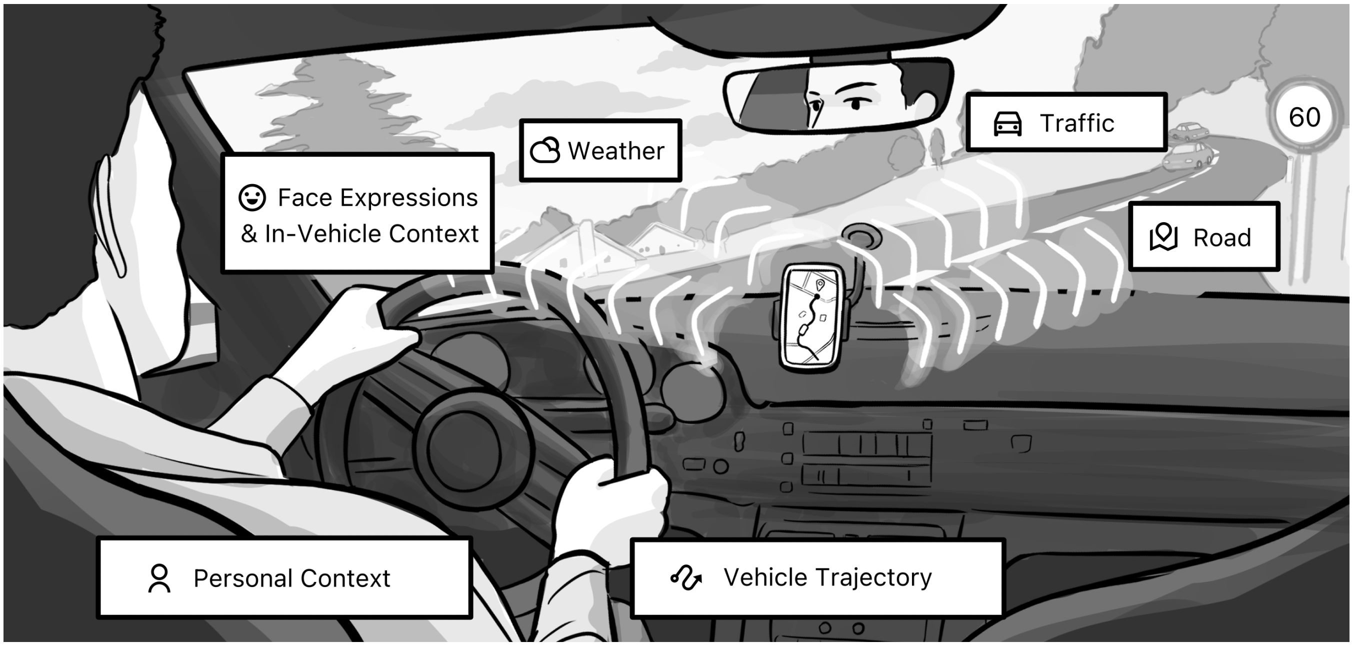


Figure 1: We present VEmotion, a new virtual emotion sensor embedded into a smartphone app that fuses an extensive variety of contextual information like vehicle- and traffic dynamics, road characterization, environmental weather, and in-vehicle context.

ABSTRACT

Detecting emotions while driving remains a challenge in Human-Computer Interaction. Current methods to estimate the driver's experienced emotions use physiological sensing (*e.g.,* skin-conductance, electroencephalography), speech, or facial expressions. However, drivers need to use wearable devices, perform explicit voice interaction, or require robust facial expressiveness. We present VEmotion (Virtual Emotion Sensor), a novel method to predict driver emotions in an unobtrusive way using contextual smartphone data. VEmotion analyzes information including traffic dynamics, environmental factors, in-vehicle context, and road characteristics to implicitly classify driver emotions. We demonstrate the applicability in a real-world driving study ($N = 12$) to evaluate

UIST '21, October 10–14, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8635-7/21/10…$15.00
https://doi.org/10.1145/3472749.3474775

the emotion prediction performance. Our results show that VEmotion outperforms facial expressions by 29% in a person-dependent classification and by 8.5% in a person-independent classification. We discuss how VEmotion enables empathic car interfaces to sense the driver's emotions and will provide *in-situ* interface adaptations on-the-go.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**; • **Computing methodologies** → *Machine learning*.

KEYWORDS

driver emotion detection, mobile sensory system, contextual affective state prediction, machine learning

ACM Reference Format:
David Bethge, Thomas Kosch, Tobias Grosse-Puppendahl, Lewis L. Chuang, Mohamed Kari, Alexander Jagaciak, and Albrecht Schmidt. 2021. VEmotion: Using Driving Context for Indirect Emotion Prediction in Real-Time. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21), October 10–14, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3472749.3474775

1 INTRODUCTION

Driving can elicit many emotional and cognitive states. The experience of driving — a combination of how one feels before entering the vehicle, the context of neighboring traffic, the behavior of other road users, the car aesthetics, and one's own driving style, among other factors — induces a wide range of emotions in drivers [48]. There is a growing interest in developing automotive user interfaces that allow for implicit and explicit interactions that are aware of how the driver is feeling [4]. This rests on the viability of the system in accurately estimating the driver's emotions, a field referred to as affective or empathic computing [7, 39].

Recent breakthroughs in ambient ubiquitous sensing [33] allow in-the-wild driver data, including real-world driving context, to inform emotion classification models. In principle, this could allow for empathic car interfaces [4] that could plan routes to invoke specific emotions, raise the user's engagement when detecting boredom by playing the user's preferred music, or mitigate undesirable driving styles that result from negative emotions (e.g., anger, sadness). The viability of such interfaces rests on the accurate, robust, and real-time classification of a driver's emotions. This remains an ongoing research challenge.

What are emotions and how do we measure them? Ekman has proposed six basic and pancultural emotions that can be inferred from one's facial features [15, 16]. This has motivated the development of computer vision for recognizing emotions from camera-captured facial expressions [2]. Besides this, implicit physiological activity could also be relied on for estimating the user's emotion. Some modalities include electroencephalography [1], electrodermal activity [8], or heart rate variability [40]. Nonetheless, physiological sensing often requires user contact with the measurement sensor, which impacts the user acceptance [25] and the overall driving experience. In comparison, remote cameras are less intrusive [31, 38]. For this reason, state-of-the-art algorithms for facial expression recognition are now commercially and widely available — such as the Affectiva SDK [34], or the Microsoft Azure face detection API[1]. These systems have been deployed on a large scale and are utilized to measure drivers' emotions and stress [9, 23]. The correlation between facial expressions and their underlying emotion can vary across individuals [45], where the emotion detection quality depends on the driver's facial expressiveness, brightness levels, and the driver's willingness to be video recorded. Here, previous research suggests that the individual driving style and driving performance are indicative of the driver's experienced emotions [22, 35]. With this in mind, we investigate whether the analysis of driving styles and driving-related information can be used to predict driver emotions? This is a counter-intuitive proposition, given that we are sensing driving information instead of sensing the driver themselves.

This paper presents VEmotion, a smartphone system that uses internal sensors only to measure driving information and estimate the perceived emotions in real-time. VEmotion analyzes the user's driving behavior through the car's surroundings variables including speed, weather, road types, and traffic flow. In contrast to previous emotion assessment modalities, VEmotion relies only on the contextual data from the vehicle that does not require modifying the car itself. To elaborate, we recorded high-dimensional contextual driving data on different routes and derived common environmental influences on emotional states. We collected data with VEmotion in a user study with twelve participants and evaluated its classification accuracy. Our results show that vehicle speed, traffic flow, and weather terms are assigned the highest feature importance from all recorded context variables. We conclude that VEmotion is an appropriate and generalizable approach for predicting the driver's emotions, achieving up to 72.4% accuracy in real-world driving scenarios.

CONTRIBUTION STATEMENT

Our work makes four contributions: (1) We present VEmotion, a mobile and personal computing software that predicts driver emotions based on contextual driving data. (2) We report an *in-the-wild* study and demonstrate that emotion recognition from camera-captured facial expressions can be improved by 28.5% using VEmotion. (3) We provide a machine learning-based processing pipeline that analyzes the relative importance of the various contextual features and, hence, their respective contribution to emotion prediction accuracy. (4) Finally, we discuss how VEmotion enables seamless emotion prediction for future empathic car interfaces. Altogether, this paper demonstrates that contextual measurements can support emotion state classification, not only of the user themselves but also of contextual variables that invoke the state (e.g., weather, traffic flow) or result from the vehicle state (e.g., car speed).

2 RELATED WORK

This section presents previous work about emotion assessment, detection of emotions in driving scenarios, and the use of emotions in interactive systems.

[1] https://azure.microsoft.com/services/cognitive-services/face, last access 2021-04-07

2.1 Emotion Assessment

There is a tendency in computer science to treat affect and emotion as the same phenomenon inferring and understanding human emotion primarily through the expression of physiological signals such as facial expression, gait, or blood conductivity [52]. Although they are different, necessary distinctions occur. Affect has been described by Deborah Gould [20] as "non-conscious and unnamed, but nonetheless registered, experiences of bodily energy and intensity that arise in response to stimuli" and thereby describes a "compound phenomenon variously consisting of evaluative, physiological, phenomenological, expressive, behavioral, and mental components" [52]. Emotion is regarded as "what from the potential of [affective] bodily intensities gets actualized or concretized in the flow of living" [20]. Treating Stark and Hoye [52] as a starting point, our current work is physiological and adopts a motivational model of emotion. We address criticism against this conflation of our chosen approach in the discussion section.

Measuring the user's emotions is a compelling topic that has been addressed by previous research. Picard coined the term *Affective Computing*, envisioning computers to express or sense emotions to provide a computerized interface that mimics human-like capabilities [39]. Modern user interfaces, such as voice or speech interfaces, benefit from understanding the user's currently perceived emotions or cognitive states and can adjust their interface according to the user's mood [54]. However, investigating robust modalities that *sense* emotions in real-time is still an ongoing research field.

Early work looked at facial expressions as a marker for perceived emotions. Ekman [14] and Ekman and Rosenberg [18] concluded that a connection between emotions and facial expressions exists. Numerous frameworks exist which can recognize emotional states using facial expressions.

However, facial expressions are considered an individual property that is different across the user's culture [43] or their gender [17]. Hence, facial expressions for interactive applications require users to calibrate towards their individual facial expressions. Kosch et al. [27] investigated if the detection of facial expressions via computer vision is feasible for mobile in-the-wild studies. They find that a re-calibration of the individual facial expressions on a per-user basis increases the correctness of emotions detected through facial expressions by 33%. However, detecting facial expressions using computer vision requires installing cameras and can compromise the user's privacy. External factors, such as illumination, can influence the quality of facial expression detection. Wearable sensors that provide a direct assessment of the user's physiological states can be used to infer the perceived emotions. Other wearable sensors exploited alternative physiological sensing modalities, such as electrodermal activity, heart rate, muscle tension, breathing rate, and electroencephalography [28]. However, wearable devices must provide a sufficient utility to the user to justify the user's effort of using the wearable sensor [59]. Also, the obtained physiological signals require a certain quality level and the suitable measurement modality for the right job to provide a meaningful assessment over the emotions [12].

2.2 Detecting Emotions while Driving

Facial expressions have a long tradition as an indicator for the expressed emotions [14]. Typical facial expressions include smiling or frowning as well as head gestures, such as nods and tilts. The detection of facial expressions requires an additional camera in the driver's cabin, including RGB cameras [9, 31, 38], infrared cameras [19] or thermal cameras [26]. Physiological sensing utilizes the driver's direct bodily responses to draw conclusions about the emotional states. Several physiological sensing modalities, such as heart rate, electrodermal activity, and electroencephalography [13, 62], are indicative of the driver's perceived emotions. However, to measure such physiological signals, sensors require direct contact with the user while driving (e.g., electrodermal activity sensor attached to the driver's hand). This can impact the driving experience and usability negatively [63]. In-car speech interfaces have been investigated as a modality to measure the driver's emotions. The way the driver talks to the voice assistant or co-drivers can indicate the user's perceived emotions. A variety of studies focus on paralinguistic features and how drivers are verbally interacting with the environment [21, 44, 46] by analyzing the sound's loudness, pitch, and spectral features [62]. However, the driver needs to communicate with an entity in the car while driving to enable robust detection of emotions which is not feasible during stressful or cognitively demanding driving scenarios.

Previous research hypothesizes that the driving behavior, style, and the driver's context are indicative of the currently perceived emotions [35]. Here, behavioral characteristics are viewed as emotional markers. For example, the grip strength applied on the steering wheel varies with the driver's emotional states [30, 36, 49]. Other factors include the interaction with the gas and brake pedals [32] as well as changes in body posture using pressure sensors [53]. Similarly, the driver's context and driving behavior are reliable factors to predict emotions. Navon et al. investigated how a driver's driving style is influenced under different emotions, finding that maladaptive driving styles are closely related with participants who have difficulties in emotion regulation and forgivingness [35]. Hancock et al. [22] show that negative emotions impact driver performance and driving styles, impacting the number of lane excursions and lateral control of the car.

Based on previous work, we expect correlations between the driving style and driver emotions. However, developers and researchers must access the car's sensor layer, which is often kept confidential, to infer the user's driving style. Standards for obtaining these data streams exist (e.g., OBD II) but are limited to specific measures, such as acceleration, braking, or steering behavior [50]. Furthermore, these standards have to be implemented by the individual car manufacturers and often miss environmental factors, including road context variables. So far, previous research has informed how emotions can be sensed in-car interfaces. Sensing the driver's emotions by utilizing the driver's driving context and behavior without modifying the user's car on the go has not been studied so far. We close this gap by presenting a study that classifies the driver's emotions by solely analyzing the context and driving behavior.

2.3 Considering Emotion Expressiveness in Real Driving Environments

Detecting emotions in the wild is a challenging task. From a machine learning perspective, most recognition models are trained with data from a constrained environment (e.g., driving simulators) and perform poorly in unconstrained scenarios. To evaluate our contribution to existing work in driver emotion recognition (e.g., with other modalities), the most recent systematic literature survey by Zepf et al. [62] provides a detailed understanding. The survey systematically reviews literature back to 2002 and identifies 63 papers on this topic. Out of 63 identified articles in the survey, only 19 papers measure emotions in natural, non-simulated settings (i.e., not induced or acted). Looking at the expressed emotion categories of the 19 papers, 16 papers were measuring stress while three papers were measuring emotions. One of these papers was predicting aggressive driving behavior without taking emotional states into account [24]. Another one used electroencephalography and electrodermal activity to predict concentration, tension, tiredness, and relaxation [41]. Finally, Riener et al. [42] inferred arousal states using electrocardiography and GPS data. Contrary to related work, our approach does not require modifying the user's car and utilizes only smartphone sensors to determine the user's driving context and behavior, hence implying the user's perceived emotions. We present the system and classification pipeline in the following section.

3 VEMOTION

In this section, we present VEmotion, a system that captures the driver's contextual driving data from the smartphone alone. We present the software architecture and the measures of our implementation in the following.

3.1 System Architecture

We implemented a smartphone app that captures contextual smartphone data to train a classifier that predicts the driver's emotions. We perform a layered approach of extracting relevant context information to learn as much as possible from the driver's driving context using a minimum set of input streams. The selected features are based on Braun et al. work [5] where driving behavior, traffic, vehicle performance, and environmental factors are relevant. We filtered the variables based on the following requirements: (1) on-device computation without accessing the vehicle itself, (2) no direct user interaction, and (3) non-critical consumption of device resources. We capture the smartphones' fused sensory data and use it as an input for a machine learning predictor. Figure 2 provides an overview of the VEmotion system architecture. VEmotion utilizes the speed of the vehicle, current weather, traffic context, road context using GPS data, and the driver's facial expressions along a perceived emotion baseline to train a predictive classifier. A prototype is developed as an iOS app, in which location-based data is sensed in a $1Hz$ (Hertz) interval, whereas the video produces

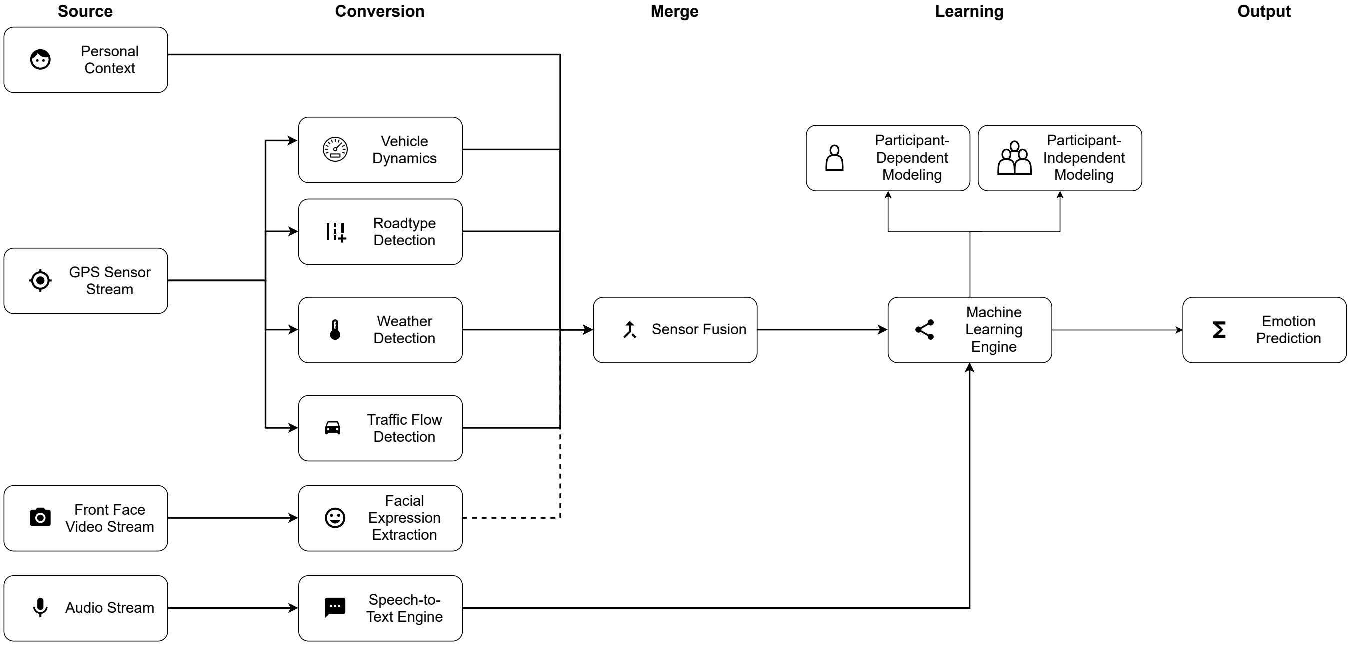


Figure 2: Overview of the VEmotion system architecture. We record contextual data (e.g., weather, road type, traffic flow) and the driver's facial expressions while driving. We fuse the collected data and use it as an input for a machine learning predictor that predicts the driver's emotions. The audio stream is used to detect the baseline emotion in our study experiment. Facial expressions can be included as a feature in VEmotion based on individual privacy policies and is therefore depicted as a dashed line. The audio stream input analyzed via a speech-text-engine is used to extract the label for our system and is not included as an input feature to the machine learning engine.

approximately 30 frames per second. In the following, we present the features and data that are recorded by VEmotion.

3.1.1 GPS Sensor: ***Vehicle Dynamics*** . We interpolate the speed of the vehicle (v) between two consecutive GPS waypoints $((lat_1, long_1), (lat_2, long_2))$ and the time between t via the Haversine formula [58]. We also calculated the vehicle's acceleration by computing the change in velocity divided by the time between using two consecutive vehicle speed measurements.

3.1.2 GPS Sensor: ***Weather***. We request weather information of each incoming GPS coordinate from the Microsoft Azure Maps API[2] to reflect the weather context conditions in real-time. Thereby, we include the following weather conditions: weather description called '*weather_term*' (e.g., '*sunny*'), the approximated outside-temperature '*feeltemp_outside*' (in $^{\circ}C$), cloud coverage '*cloud_coverage*' (in %), and wind speed '*windspeed*' (in km/h).

3.1.3 GPS Sensor: ***Trafficflow***. We also include the traffic flow in VEmotion by providing information about the speeds and travel times of the road fragment closest to the given coordinates using the Microsoft Maps Traffic Flow API. Thereby, for each GPS point we include the variable '*freeflow_speed*', which is the speed of traffic expected under ideal conditions. The freeflow speed can be different from the maximum speed limit of the road, for example, in case narrow roads force driver to slow down. To account for slow-moving traffic and jams, we define a feature called '*trafficflow_reducedspeed*'. The reduced speed of the traffic flow is calculated by the freeflow speed on the road $freeflow\_speed(lat, long)$ minus the actual traffic flow speed on this segment $current\_speed(lat, long)$: $trafficflow\_reducedspeed(lat, long) = freeflow\_speed(lat, long) - current\_speed(lat, long)$ measured in km/h.

3.1.4 GPS Sensor: ***Road Type***. We extract the nearest roads from OpenStreetMap[3] via reverse geocoding to detect the surrounding infrastructure for every GPS coordinate. We download a $200m \times 200m$ high-definition map of the current GPS coordinate and perform a map matching by calculating the euclidean distance of each node in the map to the current GPS coordinate and select the road node object that is the closest. We thereby extract the following features: road-type (e.g., '*highway*'), maximum speed on the current road (in km/h), and the number of available lanes on the current road.

3.1.5 Front-Facing Smartphone Camera: ***Facial Expressions***. We decided to include and evaluate the basic emotions captured through facial expressions [14] into our classification pipeline. The facial expression does not represent our label for predicting the emotions of the driver but is rather a way to have more inside-view information. The smartphone app obtains an image stream with 30 frames per second from the driver-facing camera and cuts it into frames to assess the driver's facial expressions. Up to 10 frames per second are sent via a cloud platform to be analyzed for facial expression features. Here, the Microsoft Face Recognition API is used to detect facial expressions that indicate specific emotions. The API returns confidences for eight basis emotions ('anger', 'contempt', 'disgust', 'fear', 'happiness', 'neutral', 'sadness', 'surprise'). No emotion value is recorded if no faces are detected (e.g., due to occlusion or shaky video stream). To have distinct emotions corresponding to a GPS coordinate rather than confidences of the eight basic emotions, we take the emotion with the maximum confidence and call this variable *facial_expression*. The validity of different cloud-based, commercial facial expression SDKs has been researched by Yang et al. [60] using a multitude of data sets such as ADFES [56], RaFD [29], WSEFEP [37]. The overall emotion recognition accuracy of Microsoft Azure is higher 84.7% compared to the 67% accuracy from the Affectiva SDK, especially 'angry', 'sad', and 'happy' facial expressions can be predicted more confidently with Microsoft Azure [60].

3.1.6 Per-Ride User-Input: ***Personal Context***. To include more subject-variant features in our analysis, we selected '*daytime*' of the ride, '*age*' of the driver, and felt emotions before the ride ('*before_emotion*') as variables to our system. Their values remain constant over the driving time.

3.2 Synchronizing Data Streams: Sensor Fusion

In the system's sensor fusion module, we make sure that all incoming sensor streams from GPS and camera are aligned along the time- and spatial dimensions. The GPS module exports its latitude and longitude signals together with the current timestamp of the sensor system in a GPX-XML format. The frontal face video stream is divided into individual frames and attached metadata about their time-occurrence based on the camera's frames per second. The output emotion categories are merged with the GPS sensor stream by the timestamp values after analyzing the individual frames and GPS-derived information. Table 1 shows the used features with example values.

Table 1: List of available features to predict emotions on the ride.

Context	Feature	Example Values
vehicle trajectory	vehicle_speed	2.255133
	vehicle_acceleration	-0.15.
weather	feeltemp_outside	13.0
	windspeed	5.6
	cloud_coverage	76
	weather_term	'clear'
traffic	trafficflow_reducedspeed	7.295495
	freeflow_speed	115.0
road	road_type	'residential'
	max_speed	30.0
	n_lanes	2
in-vehicle	facial expression	'surprise'
personal	daytime	'afternoon'
	age	21
	before_emotion	'happiness'

We performed several steps to clean the data before training a suitable context-emotion classifier. The labeling process is defined in the user study section 4. We excluded all observations, where

[2] https://azure.microsoft.com/services/azure-maps, last access 2021-04-07
[3] https://nominatim.openstreetmap.org/ui/search.html, last access 2021-04-07

the 'expressed_emotion' label is outside our specified emotion categories (e.g., one participant P11 once labeled he is 'stressed'). The string-based features are encoded into integer categories ('before_emotion', 'daytime', 'weather_term', 'road_type' and 'facial_expression') for appropriate use in the classification algorithm. Next, we cleaned the data of missing values by setting the default number of lanes 'n_lanes' to 1 and set missing entry values for 'max_speed' to 0. We selected a Random Forest Ensemble Learning as a default classifier based on a 10-fold grid-search cross-validation (using Support Vector Machines, KNeighbors, Decision Tree, Adaboost, and Random Forest classifier from scikit-learn with default parameters), in which the Random Forest achieved the highest average F_1 score. The type of modeling procedure (person-dependent and person-independent) is explained in detail in the sections 5.3 and 5.5.

We also developed a real-time prediction app of VEmotion to classify emotions on unknown roads based on the learned classifier in which the mean emotion inference took 1.36s (SD: 0.246, min: 0.962, max: 1.996) in a 30-minute test ride.

4 USER STUDY

We conduct a user study to understand the impact of the VEmotion's contextual data on emotion prediction.

4.1 Apparatus and Method

We built a vehicle-usable iOS app that records the individual GPS and video stream and computes the variables described in Section 3 continuously during the ride[4]. We asked the participants to use this app the next time they used their personal car to ride and attach their phone to the windscreen. We recorded the daytime and asked the participant about their currently perceived emotion at the beginning of the ride. To collect a baseline of the participant's own interpretation of emotional states during the ride, we trigger a beep tone every 60 seconds for the participants to verbally provide their currently perceived emotions. We designed this emotion probing in correspondence to the *in-situ* categorical emotion response (CER) rating for collecting data on emotional experiences in vehicles [11]. Participants were instructed about the set of available emotions before starting the experiment (i.e., the basic emotions after Ekman [14]). The verbally expressed emotion was recorded while driving and is analyzed after the driving scenarios with a speech-to-text algorithm. As this procedure requires the passenger to talk during the ride and can be a distraction from first-order driving tasks, in a pre-study ($N = 5$) we optimized the time interval not to be annoying, ensure safety, and simultaneously cover the felt emotions appropriately. A post-hoc driving questionnaire showed that 9/12 participants were not bothered by the beep. The mean time-to-beep-response was 1.8 seconds. For an in-the-wild system that uses our architecture, the ground truth emotion assessment will not be required, and therefore, the system will not interact with the driver. A printout of the basic emotions was given to the participants before the start of the experiment. After the ride, the participant answers several subjective questions, including remarkable incidents.

4.2 Procedure

Twelve participants were invited through a mailing list from a pool of colleagues willing to participate in research studies. They were asked to download our iOS app beforehand and were equipped with a windshield smartphone retainer. The participants were asked before their next ride to call the study instructor via a remote call. In this call, the participants were asked about their demographics, frequency of driving, and feelings before the ride. Then we gave an introduction to our app. The participants were then asked to hang up, start the app recording, and drive freely to their chosen destination and after the ride to save the recordings and call the instructor. The instructor asked the participants about notable incidents while driving, emotions while and after driving.

4.3 Participants

We recruited 12 participants (eight self-identified as male, two self-identified as female) with an average age of 27 years (SD = 4.73). Six participants occasionally drive (i.e., less than 10,000 kilometers per year), where three participants drive moderate distances (i.e., between 10,000 and 20,000 kilometers per year), and three participants drive more frequently (i.e., more than 20,000 kilometers per year). The mean duration of the rides is 16 minutes (SD = 11, min=7, max=52). The road type changed on average 7.9 times per ride. Participants expressed on average 4.41 distinct emotions during their ride (the duration between different expressed emotions across all users was 2 minutes 43 (SD=3 minutes 59).

5 RESULTS

We analyze the prediction performance of driver's emotions using the data captured by VEmotion. First, we evaluate the relative importance of single features of the data set collected by VEmotion. Then, we investigate the classification accuracy for emotion recognition based on facial expressions alone. Finally, we performed the following model evaluations: (1) a Leave-One-of-10-Road-Segments-Out cross-validation, (2) a participant-dependent Leave-One-of-10-Road-Segments-Out cross-validation, and (3) a Leave-One-Participant-Out cross-validation for evaluating VEmotion on unseen participants (i.e., participant-independent evaluation).

5.1 Relevant Features for Predicting Emotions

We collected 8986 instances of labeled data, namely a GPS location with a ground-truth label of the user's self-reported emotion. This corresponds to 1.1 seconds of driving depending on data validity, such as GPS fixes. Overall, 5780 were labeled as 'neutral' (64%), 2839 as 'happy' (32%), 177 as 'surprise' (2%), 130 as 'angry'(1%), and 60 as 'disgust'(< 1%). We start by investigating how decisive each feature was for creating a classification model. For this, we extracted the feature importance of the context variables, provided by VEmotion, in a leave-one-participant-out situation in Figure 3. As we employed a Random Forest classifier for emotion prediction, feature importance is measured as the popular mean decrease in impurity — this is defined as the total decrease in node Gini-impurity (weighted by the probability of reaching that node), averaged over all trees of the ensemble [6].

Of the context variables, 'vehicle_speed' was ranked highest in terms of feature importance. This might be because 'happy'

[4] Ethical approval was granted by the institutional review board of the university department

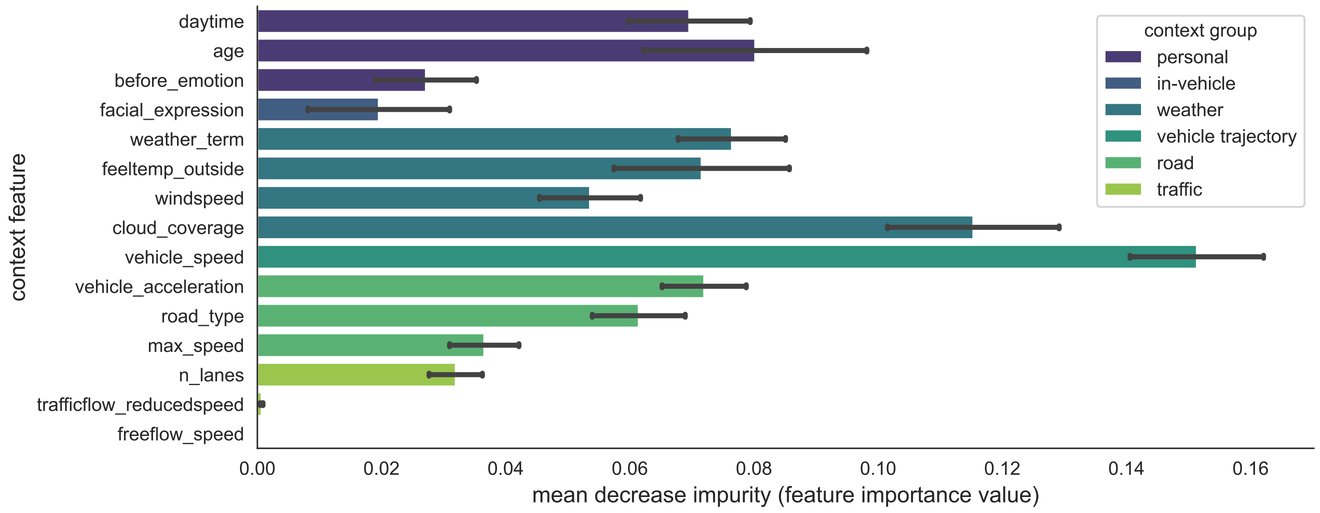


Figure 3: Feature importances measured by the mean decrease of Gini-impurity for the Leave-One-Participant-Out cross-validation.

emotions are often reported in unhindered speed scenarios. In contrast, related research [61] report higher negative emotions (i.e., 'anger' and 'fear') during unforeseen traffic incidents (e.g., high traffic densities or red light series) that require high cognitive demands. The information extracted by the traffic variables ('trafficflow_reducedspeed' and 'freeflow_speed') is assigned the lowest feature importances overall, which might be due to the vehicle trajectory features (acceleration and speed) working as proxy variables for unhindered traffic rides. Interestingly 'weather' and 'daytime' were assigned a medium level of feature importance. These environmental variables have been observed to impact emotional states in related psychological research [10, 61]. Related research has weakly associated negative emotional states to 'temperature' and positive emotions to 'sunlight'. However, weather influences tend to be highly dependent on person and age, which are additional context variables in VEmotion. Contrary to our expectations, the emotion reported before the ride was not assigned a very high feature importance, which may be due to mood changes when driving and unforeseen traffic incidents. The facial expressions captured by the frontal face camera have low feature importance. In contrast, all other recorded context inputs have medium-level importance. This underlines the usefulness of personal- and environmental input based on GPS location.

In a subsequent analysis, we evaluated the learned feature importances assigned conditional to the emotional class labels. We observe that 'cloud_coverage' and 'max_speed' information contribute highly to 'happy' emotions. Interestingly, 'freeflow_speed' has high feature importance conditioned on 'disgust' emotional states.

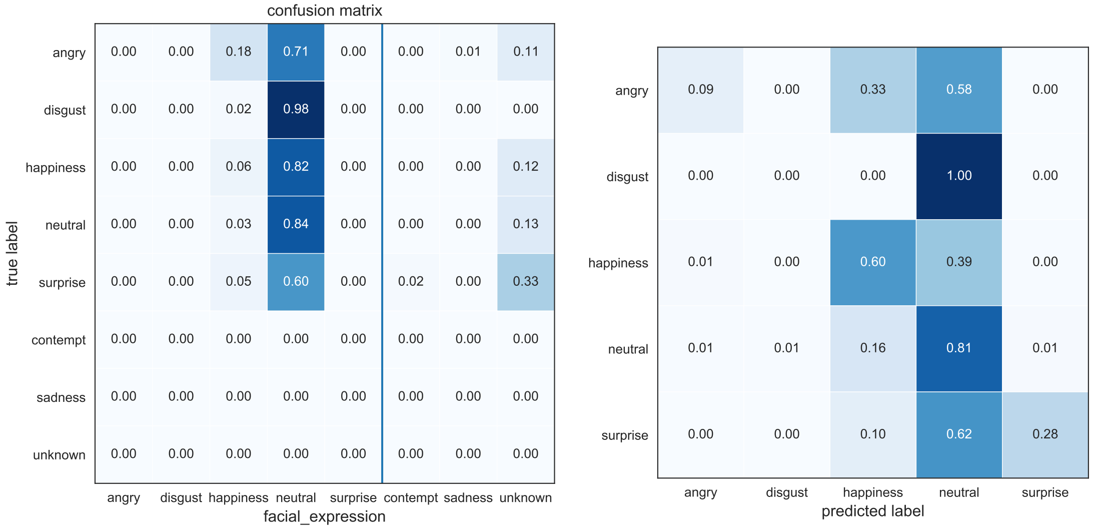
5.2 Validity of Facial Expressions

We analyzed the validity of 'facial expressions' as the sole indicator for the driver's emotions. We observe that 'facial expressions' predict five distinct emotional categories on the complete data set ('neutral', 'happiness', 'surprise', 'contempt', 'sadness', and 'unknown' if no face is detected). The emotions reported by participants on the ride are 'angry', 'disgust', 'happiness', 'neutral', and 'surprise', and thereby a subset of the facial expressions detected. Figure 4a shows the output of the facial expression engine and the self-reported emotion in our data set. The confusion matrix indicates that the facial expression engine detects many 'neutral' emotions (which are, in most cases, the true self-reported emotions). However, the self-reported emotion is often not correctly detected: 60% of 'surprised' emotions are predicted as 'neutral' emotions. At the same time, 82% of 'happy' and 98% of 'disgust' emotions are predicted as being 'neutral'. To conclude, the facial expression engine often yields a 'neutral' emotion class, ignoring and misclassifying heavily other felt emotions of the driver.

5.3 Leave-One-of-10-Road-Segments-Out Cross-Validation

Emotion recognition from 'facial expressions' alone is limited. To overcome this, we trained a Random Forest classifier (random state = 0, n_estimators = 50, max_features= $log2$[5]) on the whole data set in the study, which included 'context variables'. We performed unshuffled cross-validation with 10-folds from all participants by segmenting the participant data into ten distinct consecutive folds (time-dependent road segments). Thereby, we construct a training set from nine training folds and one test set consisting of the remaining folds. This avoids a common constraint posted by a traditional 10-fold shuffled cross-validation evaluation since neighboring samples can be present in both training and test sets, resulting in trivial classification models. We term our evaluation approach 'Leave-One-of-10-Road-Segments-Out cross-validation', as this provides a better picture of the potential performance and robustness for

[5] The hyperparameters are found using a 10-fold hyperparameter tuning grid search.



(a)

(b)

Figure 4: Comparison between facial expression output vs. VEmotion in predicting self-reported emotions on the road. The values of the confusion matrices are normalized on the true emotion class occurrences. The confusion matrix have different sizes as the facial expression engine falsely outputs a larger set of emotions (indicated by the blue vertical line). (a): Direct output of the facial expressions from Microsoft Azure. The detection accuracy of the self-reported emotions is 55.57%. **(b): 10-fold cross-validation on the participants' unshuffled data using VEmotion accessing contextual data (no facial expression features included) trained with a Random Forest classifier trained with an accuracy of** 71.70%. **VEmotion achieves an unweighted average of the class specific recalls of 0.41 in (a) vs. a worse close-to-chance 0.18 when using facial expressions alone (b).**

evaluating the classification performance. Furthermore, we believe that realistic data sets will contain relatively few, hence insufficient, 'angry' and 'disgust' emotion categories for model learning. Due to the imbalance of emotional classes that may be apparent in specific rides, we set the class weight of observations to 'balance'. This means that the Random Forest in VEmotion uses the values of the emotion class to automatically adjust weights inversely proportional to class frequencies in the training folds. Hence, we calculate the weighted average F_1 score over all emotion classes, which is defined as the harmonic mean of precision and recall, as an evaluation measure of classification performance.

VEmotion prediction performance of self-reported emotions in a 10-fold cross-validation on unseen ride segments is shown in Figure 4b. The overall accuracy of emotions is 71.70%. In other words, it is 29% better than relying on the 'facial expression' engine alone. We validated the facial expression engine by using other common facial expression classifier systems. We explored and applied a locally computable EmoPy trained on FER 2013 dataset [55] and AWS Emotion Recognition [47] to our data showing similar, subpar results (predicting neutral/calm states is prevalent, accuracy: 0.55 and 0.07). VEmotion achieves an weighted average F_1 score of 71.30 (SD: 0.0713) across all emotional classes and outperforms the facial-expression-only system by 20 percentage points. We also observe that VEmotion only predicts classes that are actually expressed during the ride. In contrast, the 'facial expression' engine predicted contempt' or 'sad' emotions. Furthermore, VEmotion predicts 60% of 'happy' emotions vs. 6% using facial expressions only by only losing 3%. of correct 'neutral' emotion predictions. 'surprise' emotions can be accurately predicted with 28%. In contrast, 'angry' and 'disgust' emotions cannot be properly detected by VEmotion. The results indicate that contextual information can significantly improve the classification of emotional states, especially in detecting 'surprise' situations. VEmotion additionally discriminates better between 'neutral' and 'happiness' states of the driver. This evaluation is based on a 10-fold cross-validation and has access to training data of individual participants. We show that we can learn a global system for recognizing emotions 'on-the-go' with contextual and facial expressions. However, this comes at higher computational costs of having access to all participants' data and learning a participant-independent classifier. If the system should be used for uncalibrated modeling of a new driver's emotions, we perform an extensive evaluation in the next paragraph.

5.4 Participant-Dependent Leave-One-of-10-Road-Segments-Out Cross-Validation

Furthermore, we analyzed participant-dependent modeling using a participant-dependent Leave-One-of-10-Road-Segments-Out cross-validation. This means that we are training a participant-dependent model and validating on a holdout set of the participant using a 10-fold cross-validation scheme. The results are presented in Table 2.

Table 2: Accuracy, precision, recall, and weighted F_1 scores of the global 10-fold cross validation on unseen consecutive driving segments, aggregate of participant-dependent Leave-One-of-10-Road-Segments-Out cross validation, as well as leave-one-participant-out cross-validation.

	Leave-One-of-10-Road-Segments-Out Cross-Validation				**Participant-Dependent Leave-One-of-10-Road-Segments-Out Cross-Validation**				**Leave-One-Participant-Out Cross-Validation**			
Input	Accuracy	Precision	Recall	F_1	Accuracy	Precision	Recall	F_1	Accuracy	Precision	Recall	F_1
Facial Expressions	.56	.57	.56	.51	.57	.66	.57	.56	.59	**.63**	.59	.54
VEmotion + Facial Expressions	.72	1.0	.72	**.72**	**.70**	.86	.70	.73	.64	.58	.64	**.57**
VEmotion	**.72**	**1.0**	**.72**	.71	**.71**	**.89**	**.71**	**.73**	**.64**	.56	**.64**	.56

5.5 Leave-One-Participant-Out Cross-Validation

We evaluate the possibility of a general classification model using all participant data except for one for training and using the last participant for evaluation. Semantically, this approach learns a model without knowing anything about the driver in advance and predicts the drivers' emotions independent from individual context emotion preferences. As we have a more complex prediction problem by not having learned from the held-out participant, we expect the overall prediction to decrease. The results of the experiment are shown in Table 2.

5.6 Comparison of Model Performances

Table 2 provides an overview of the prediction performance scores for the different evaluation procedures based on VEmotion. The Leave-One-of-10-Road-Segments-Out cross-validation approach, which uses all data samples from all participants, yields 71.70% accuracy. This is considerably higher than relying on 'facial expressions' only, which achieves a mean accuracy of 55.58%. Looking at the F_1 score, weighted for the class labels, VEmotion achieves a score of 0.7130 and 0.7164 with inclusion of facial expressions. In the next step, we performed a 10-fold cross validation only based on individual participants' data and aggregate the results over all participants (i.e., a participant-dependent Leave-One-of-10-Road-Segments-Out cross-validation). Here, we observe a similar prediction performance compared to the global model. VEmotion achieves here an average accuracy of 70.67% which is approximate 1.03%. smaller than in the global 10-fold cross-validation step. Also, the weighted F_1 score increased slightly to 0.7282.

In the participant-dependent cross-validation, we also observe that the VEmotion without the variables from the 'facial expression' (VEmotion) engine has a marginally higher precision of 88.59 % than VEmotion including facial expressions. This indicates that a high fraction of emotions are predicted with a low false-positive fraction. Looking at a much more challenging problem of predicting emotion categories of unseen participants in the Leave-One-Participant-Out scheme, the virtual sensor also outperforms the other models with an F_1 score of 0.56. The average accuracy of VEmotion is 63.71%. This is less than the achieved accuracy in the 10-fold cross-validation but remains a high-quality predictor if no information about the user is known in this multiple class output prediction. Since global and participant-dependent modeling of contextual emotions yield similar prediction qualities, we conclude that it is computationally favorable to learn participant-wise models over various rides, instead of learning global models that require data exchange of all participants. This also ensures that the inter-person and trip variety is sufficiently accounted for in the training sample. We stress the fact of imbalanced emotion classes that can only be acquired mainly through global data acquisition. Thereby learning a solely participant-dependent model puts the detection of emotions at a disadvantage that are not occurring frequently (e.g., 'surprise', 'angry', 'fear', and 'disgust'). Hence, facial expressions that do not occur frequently can still contribute to a robust model when collecting them from multiple participants. Finally, our 'in-the-wild study' does not show a significant benefit in including 'facial expressions' as features in our classification model. Thus, we propose omitting 'facial expressions' in practice, which would further reduce computational costs. Furthermore, facial expressions inhibit largely privacy concerns of the end-users and might raise a feeling of video surveillance while driving.

To answer the question of how many minutes of driving data is needed for VEmotion to be accordingly calibrated to predict emotions on the road with high accuracy,

5.7 Participant Fine-Tuning

We added a learning scheme below that illustrates how many minutes of driving data is needed for VEmotion to be calibrated for a high accuracy emotion prediction on the road. We used a leave-one-participant-out classifier to assess the emotion classification performance by incorporating the first x minutes of additional participants' driving data and evaluated the performance of VEmotion on the remaining driving data. Figure 5 presents the results of the analysis. We found that the first five to ten minutes have to be captured to achieve a mean precision of over 75% across participants due to the better discrimination of the classifier between neutral and happy states during the first 10 minutes (F_1 = .61). The drop in accuracy and F_1 after 10 minutes of training data is due to little held out test data which increases the variability of the prediction intervals. High precision of 80% and recall of 63.5% can be achieved when fine-tuning the classifier on the first 14 minutes. However, this requires the driver to label his perceived emotions 14 times, which may be annoying if done on every ride. We suspect our performances to increase heavily if multiple rides with fine-tuning in the first minutes are performed. Furthermore, VEmotion's application in practice would benefit from perceived emotion labeling in any special scenario within the ride and not just during the first minutes of driving.

6 DISCUSSION

Can we predict driver emotions based on driving context? To the best of our knowledge, VEmotion provides the first in-car sensor that combines implicit and non-intrusive measures to detect the driver's emotional states. In a user study with twelve participants, we find the highest classification accuracy when training a global model. We discuss the implications of our results in the following.

6.1 Driving Context Implies Emotions

Previous work hypothesized that the observation of driving behavior can be indicative of driver emotions [22, 35]. Indeed, our results show that certain features are predictive for driver emotions. In analyzing the feature importances, we found that 'vehicle dynamics', 'weather', and 'traffic flow' were highly predictive for emotions. This implies that the designer of empathic car interfaces should focus on the reliable measurement of these features when assessing emotions is a critical task. These can be integrated into existing emotion recognition engines or car navigation systems that are already integrated into vehicles or smartphones. Our results demonstrate that different users share common emotional categories influenced by the same contextual and environmental factors. In our real-world study, we notice a high imbalance of self-reported emotions, as most people either respond to feeling 'happy' or 'neutral' along their ride. This provides a challenging task for an appropriate data basis and proper classification of 'sad', 'fear', or 'disgust' states which are often observed with higher safety concerns [62]. These imbalanced emotion class distributions in the wild should therefore be extended in future data acquisition.

6.2 Comparing the Classification Performance between Facial Expressions and Driving Behavior

We find a difference between the classification performance for VEmotion, facial expressions, and a combination of VEmotion and facial expressions. Our study shows that the use of facial expressions alone results in the lowest classification accuracy compared to either VEmotion or VEmotion in combination with facial expressions in a real-world driving setting. Furthermore, our results show that not all emotions can be reliably detected using facial expressions. This includes the emotions 'angry', 'surprise', or 'disgust'.

Our results show that the emotion class 'neutral' is predicted most often by the facial expression engine. We suspect that the 'neutral' emotion class occurs frequently due to the low facial expressiveness in driving scenarios. Also, facial expressions are affected by user-to-user variability, resulting in individual differences in facial expressiveness and self-reported emotions. Further limitations include a moving driving environment, occlusion, and changing visibility conditions (e.g., sudden darkness in a tunnel). In contrast, VEmotion captures the driving behavior of the user in addition to facial expressions, which introduced performance increases of 38% in person-dependent (Leave-One-of-10-Road-Segments-Out cross-validation) and 10% in person-independent cross-validation schemes.

While our results show an improved classification performance for VEmotion, we find that the driving behavior and the perceived emotions are individual factors. Here, the resulting general model results in poor classification performances. However, training the model for each user individually yields a higher classification accuracy. VEmotion has to learn person-dependent discriminatory features from the contextual data to achieve acceptable accuracies. Therefore, the emotions predicted by VEmotion improves if more person-dependent information is available.

6.3 Enabling Empathic Vehicle-Applications with VEmotion

VEmotion allows the implementation of several use cases, however, our work intends to make a sensory system contribution of unobtrusively measuring emotions in the wild. VEmotion is beneficial in providing direction into what enjoyable drives are, and VEmotion's predictions[6] can inform infrastructure and road planning policies. For instance, it might be meaningful to enforce speed limits or narrow roads on some road segments to increase the overall road safety based on VEmotion. For example, VEmotion enables navigation functionality to invoke positive emotions. This idea has been proposed but yet has to be implemented [4]. Unknown route segments can be labeled with the respectively measured emotions. Car navigation can then be extended by routing after emotions. Other applications include the reflection of emotions after a ride. For example, a post-driving tool can visualize the perceived emotions for single road segments. Furthermore, future empathic car interfaces can utilize VEmotion to modulate driver emotions in real-time, for example, by playing pleasurable music [57].

6.4 Ethical Considerations

We emphasize an ethical as well as transparent use of VEmotion for application purposes and stress that emotions are intimate, personal, and vulnerable, where potential emotional insights can be manipulated to impact behavior in the long term [3]. Until now, many resources went into in-vehicle sensing which has resulted in much debate about the need for limiting facial recognition technology due to privacy and ethical considerations [3, 51]. The current work objectively looks at the significance of facial recognition and other data regarding what they might be telling us about the human perceived emotion. To the best of our knowledge, this is the first study where volunteers have allowed the recording of facial expressions together with contextual vehicular data in the wild. Our analysis reveals that contextual data obtained from a vehicle-CAN or smartphone is more efficient than actual facial recognition technologies. This has implications on several fronts: (1) We have been collecting vehicle data for the last 15 years, yet a potential exploit of this data might enable to backwards-infer human's perceived

[6] given a more broaden data acquisition

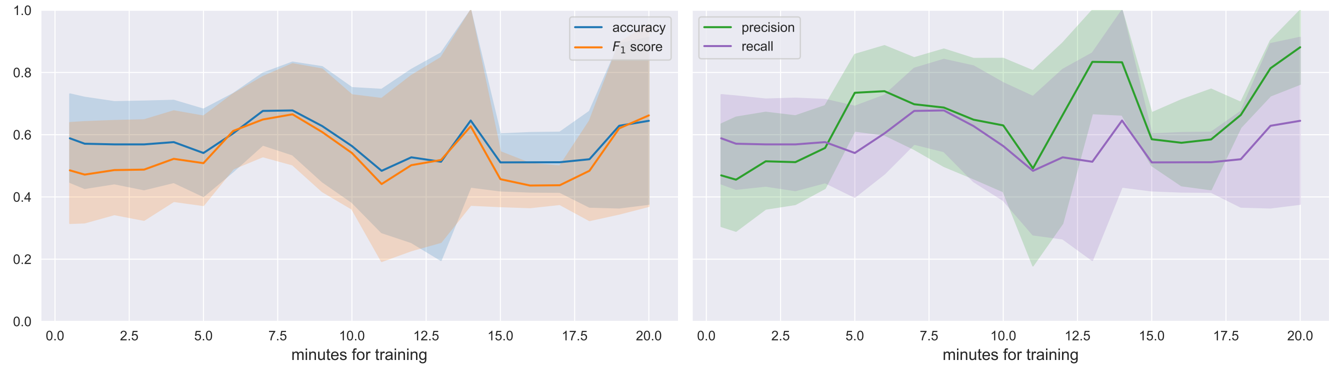


Figure 5: Classification performance for fine-tuning the Random Forest classifier with the first x minutes of participants riding data in a Leave-One-Participant-Out training scheme, in which the performance is computed on the available rest-duration of the ride data. The mean performance across all participants, as well as the 95% confidence intervals, are shown. Overall the performance of the classifier increases in all metrics (accuracy, F_1 score, precision and recall) knowing the first 5 minutes of personal driving data. Subsequently, the metrics converge, but precision is steadily increasing. We stopped computing the performance after 20 minutes due to little held out remaining driving test data.

feeling on this road given the features presented are available in the data. (2) Environmental contextual data offers a potentially more privacy-preserving and discomfort-reducing alternative to measure emotions in the wild. The connection between affect and emotions has always been emphasized. However, many other data variables can infer emotions without the need for recording affective or physiological variables. Our current work broadens the debate as to what type of data should be accessible by whom and for what purposes.

6.5 Limitations and Future Work

The robustness of our approach relies heavily on the quality of contextual input sensors. Thus, reliable in-vehicle emotion classification becomes less reliable as more features drop out. For example, facial expressions require a particular “expressiveness” of the driver to detect the emotion. Another example includes the dropout of contextual driving data, such as GPS connectivity, when driving through a tunnel. We also do not gain introspective insights on the on-goings of the driver’s mind and instead describe the driver’s perceived emotions via eight primary states; this abstracts a significant part in the much broader assessments of the multitude of felt psychological on-goings of the driver. To further reflect the relationship between emotional contextual triggers and emotional states, we will expand our work to include outside-view-camera input. Expressions via ecstatic hand gestures indicating angry affective states could not be found in the video stream but may also provide a direction for future camera-based affect features. Future work might also extend the outside-view of VEmotion by looking at other car’s behavior through the use of more privacy concerning frontal video stream input. A more extensive database of rides with a wider variety and distinction of emotions and more extended personal driving history enables longitudinal studies. Here, we strongly stress acknowledging the context of the driver and surroundings besides the emotion prediction, which should be represented in the decision space of empathic car interfaces and data basis for emotion recognition engines. Finally, our results show that an 8-minute calibration procedure on unseen drivers is sufficient to achieve a satisfying accuracy of over 68%, while the beep sound was not perceived as annoying by the participant. However, different unobtrusive strategies for a suitable calibration of VEmotion, such as incident-based sampling, will be evaluated in future work, having the caveat of not accessing a high-resolution emotion assessment on all road types.

7 CONCLUSION

This paper presents VEmotion, a system that derives user emotions by assessing driving information. We found that context variables can be captured in real-time using GPS at low cost, optionally accompanied by a camera monitoring the driver. This finding is unique as comparatively few studies are performed ‘in-the-wild’ and with the use of personal computing devices as opposed to the bespoke in-vehicle sensors. We gain many insights by having the ability to record a much more fine-grained picture of the driver and its surroundings and potential influences on emotion with VEmotion in a noisy real-world environment. This provides automotive user interface designers with an additional tool to design unobtrusive empathic car interfaces deployed in real-world scenarios. Here, we are confident that VEmotion advances the field of emotion-aware car interfaces. To encourage research in this area, we publish the source code of VEmotion and the data set for further analysis by the research community[7].

ACKNOWLEDGMENTS

This research was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in TRR161 (Quantitative methods for visual computing, Project ID 251654672) in Project C06.

[7] https://github.com/davebeght/VEmotion

REFERENCES

[1] S. M. Alarcão and M. J. Fonseca. 2019. Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing* 10, 3 (2019), 374–393. https://doi.org/10.1109/TAFFC.2017.2714671

[2] Victor M Álvarez, Claudia N Sánchez, Sebastián Gutiérrez, Julieta Domínguez-Soberanes, and Ramiro Velázquez. 2018. Facial emotion recognition: a comparison of different landmark-based classifiers. In *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*. IEEE, 1–4.

[3] Nazanin Andalibi and Justin Buss. 2020. The human in emotion recognition on social media: Attitudes, outcomes, risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[4] Michael Braun, Jingyi Li, Florian Weber, Bastian Pfleging, Andreas Butz, and Florian Alt. 2020. What If Your Car Would Care? Exploring Use Cases For Affective Automotive User Interfaces. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 37, 12 pages. https://doi.org/10.1145/3379503.3403530

[5] Michael Braun, Florian Weber, and Florian Alt. [n.d.]. Affective Automotive User Interfaces - Reviewing the State of Emotion Regulation in the Car. In *To appear in ACM Copmuting Surveys*.

[6] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[7] Yang Cai. 2006. Empathic computing. In *Ambient Intelligence in Everyday Life*. Springer, 67–85.

[8] Delphine Caruelle, Anders Gustafsson, Poja Shams, and Line Lervik-Olsen. 2019. The use of electrodermal activity (EDA) measurement to understand consumer emotions – A literature review and a call for action. *Journal of Business Research* 104 (2019), 146–160. https://doi.org/10.1016/j.jbusres.2019.06.041

[9] Silvia Ceccacci, Maura Mengoni, Generosi Andrea, Luca Giraldi, Giuseppe Carbonara, Andrea Castellano, and Roberto Montanari. 2020. A Preliminary Investigation Towards the Application of Facial Expression Analysis to Enable an Emotion-Aware Car Interface. In *Universal Access in Human-Computer Interaction. Applications and Practice*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 504–517. https://doi.org/10.1007/978-3-030-49108-6_36

[10] Marie Connolly. 2013. Some like it mild and not too wet: The influence of weather on subjective well-being. *Journal of Happiness Studies* 14, 2 (2013), 457–473. https://doi.org/10.1007/s10902-012-9338-2

[11] Monique Dittrich and Sebastian Zepf. 2019. Exploring the validity of methods to track emotions behind the wheel. In *International Conference on Persuasive Technology*. Springer, 115–127. https://doi.org/10.1007/978-3-030-17287-9_10

[12] Maria Egger, Matthias Ley, and Sten Hanke. 2019. Emotion Recognition from Physiological Signal Analysis: A Review. *Electronic Notes in Theoretical Computer Science* 343 (2019), 35–55. https://doi.org/10.1016/j.entcs.2019.04.009 The proceedings of AmI, the 2018 European Conference on Ambient Intelligence.

[13] Maria Egger, Matthias Ley, and Sten Hanke. 2019. Emotion recognition from physiological signal analysis: a review. *Electronic Notes in Theoretical Computer Science* 343 (2019), 35–55. https://doi.org/10.1016/j.entcs.2019.04.009

[14] Paul Ekman. 1984. Expression and the nature of emotion. *Approaches to emotion* 3, 19 (1984), 344.

[15] Paul Ekman. 1992. Are there basic emotions? (1992). https://doi.org/10.1037/0033-295X.99.3.550

[16] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384. https://doi.org/10.1037/0003-066X.48.4.384

[17] Paul Ekman, Wallace V Friesen, Maureen O'sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E Ricci-Bitti, et al. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology* 53, 4 (1987), 712. https://doi.org/10.1037/0022-3514.53.4.712

[18] Rosenberg Ekman. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

[19] H. Gao, A. Yüce, and J. Thiran. 2014. Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE International Conference on Image Processing (ICIP)*. 5961–5965. https://doi.org/10.1109/ICIP.2014.7026203

[20] Deborah Gould. 2010. On affect and protest. In *Political emotions*. Routledge, 32–58.

[21] Michael Grimm, Kristian Kroschel, Helen Harris, Clifford Nass, Björn Schuller, Gerhard Rigoll, and Tobias Moosmayr. 2007. On the Necessity and Feasibility of Detecting a Driver's Emotional State While Driving. In *Affective Computing and Intelligent Interaction*, Ana C. R. Paiva, Rui Prada, and Rosalind W. Picard (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 126–138. https://doi.org/10.1007/978-3-540-74889-2_12

[22] GM Hancock, PA Hancock, and CM Janelle. 2012. The impact of emotions and predominant emotion regulation technique on driving performance. *Work* 41, Supplement 1 (2012), 3608–3611. https://doi.org/10.3233/WOR-2012-0666-3608

[23] Javier Hernandez, Daniel McDuff, Xavier Benavides, Judith Amores, Pattie Maes, and Rosalind Picard. 2014. AutoEmotive: Bringing Empathy to the Driving Experience to Manage Stress. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS Companion '14)*. Association for Computing Machinery, New York, NY, USA, 53–56. https://doi.org/10.1145/2598784.2602780

[24] Ozgur Karaduman, Haluk Eren, Hasan Kurum, and Mehmet Celenk. 2013. An effective variable selection algorithm for Aggressive/Calm Driving detection via CAN bus. In *2013 International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, 586–591. https://doi.org/10.1109/ICCVE.2013.6799859

[25] Armağan Karahanoğlu and Çiğdem Erbuğ. 2011. Perceived Qualities of Smart Wearables: Determinants of User Acceptance. In *Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces* (Milano, Italy) *(DPPI '11)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/2347504.2347533

[26] A. Kolli, A. Fasih, F. A. Machot, and K. Kyamakya. 2011. Non-intrusive car driver's emotion recognition using thermal camera. In *Proceedings of the Joint INDS'11 ISTET'11*. 1–5. https://doi.org/10.1109/INDS.2011.6024802

[27] Thomas Kosch, Mariam Hassib, Robin Reutter, and Florian Alt. 2020. Emotions on the Go: Mobile Emotion Assessment in Real-Time Using Facial Expressions. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 18, 9 pages. https://doi.org/10.1145/3399715.3399928

[28] Janina Künecke, Andrea Hildebrandt, Guillermo Recio, Werner Sommer, and Oliver Wilhelm. 2014. Facial EMG responses to emotional expressions are related to emotion perception ability. *PloS one* 9, 1 (2014), e84053. https://doi.org/10.1371/journal.pone.0084053

[29] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, Skyler T Hawk, and AD Van Knippenberg. 2010. Presentation and validation of the Radboud Faces Database. *Cognition and emotion* 24, 8 (2010), 1377–1388.

[30] Y. Lin, H. Leng, G. Yang, and H. Cai. 2007. An Intelligent Noninvasive Sensor for Driver Pulse Wave Measurement. *IEEE Sensors Journal* 7, 5 (2007), 790–799. https://doi.org/10.1109/JSEN.2007.894923

[31] Zhiyi Ma, Marwa Mahmoud, Peter Robinson, Eduardo Dias, and Lee Skrypchuk. 2017. Automatic Detection of a Driver's Complex Mental States. In *Computational Science and Its Applications*, Osvaldo Gervasi, Beniamino Murgante, Sanjay Misra, Giuseppe Borruso, Carmelo M. Torre, Ana Maria A.C. Rocha, David Taniar, Bernady O. Apduhan, Elena Stankova, and Alfredo Cuzzocrea (Eds.). Springer International Publishing, Cham, 678–691. https://doi.org/10.1007/978-3-319-62398-6_48

[32] L. Malta, C. Miyajima, N. Kitaoka, and K. Takeda. 2011. Analysis of Real-World Driver's Frustration. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (2011), 109–118. https://doi.org/10.1109/TITS.2010.2070839

[33] E. Massaro, C. Ahn, C. Ratti, P. Santi, R. Stahlmann, A. Lamprecht, M. Roehder, and M. Huber. 2017. The Car as an Ambient Sensing Platform [Point of View]. *Proc. IEEE* 105, 1 (2017), 3–7. https://doi.org/10.1109/JPROC.2016.2634938

[34] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 3723–3726. https://doi.org/10.1145/2851581.2890247

[35] Meital Navon and Orit Taubman – Ben-Ari. 2019. Driven by emotions: The association between emotion regulation, forgivingness, and driving styles. *Transportation Research Part F: Traffic Psychology and Behaviour* 65 (2019), 1–9. https://doi.org/10.1016/j.trf.2019.07.005

[36] Michael Oehl, Felix W. Siebert, Tessa-Karina Tews, Rainer Höger, and Hans-Rüdiger Pfister. 2011. Improving Human-Machine Interaction – A Non Invasive Approach to Detect Emotions in Car Drivers. In *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments*, Julie A. Jacko (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 577–585.

[37] Michal Olszanowski, Grzegorz Pochwatko, Krzysztof Kuklinski, Michal Scibor-Rylski, Peter Lewinski, and Rafal K Ohme. 2015. Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in psychology* 5 (2015), 1516. https://doi.org/10.3389/fpsyg.2014.01516

[38] M. Paschero, G. Del Vescovo, L. Benucci, A. Rizzi, M. Santello, G. Fabbri, and F. M. F. Mascioli. 2012. A real time classifier for emotion and stress recognition in a vehicle driver. In *2012 IEEE International Symposium on Industrial Electronics*. 1690–1695. https://doi.org/10.1109/ISIE.2012.6237345

[39] Rosalind W Picard. 2000. *Affective computing*. MIT press.

[40] Daniel S. Quintana, Adam J. Guastella, Tim Outhred, Ian B. Hickie, and Andrew H. Kemp. 2012. Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition. *International Journal of Psychophysiology* 86, 2 (2012), 168–172. https://doi.org/10.1016/j.ijpsycho.2012.08.012

[41] Genaro Rebolledo-Mendez, Angelica Reyes, Sebastian Paszkowicz, Mari Carmen Domingo, and Lee Skrypchuk. 2014. Developing a body sensor network to detect emotions during driving. *IEEE transactions on intelligent transportation systems* 15, 4 (2014), 1850–1854. https://doi.org/10.1109/TITS.2014.2335151

[42] Andreas Riener, Alois Ferscha, and Mohamed Aly. 2009. Heart on the road: HRV analysis for monitoring a driver's affective state. In *Proceedings of the 1st*

international conference on automotive user interfaces and interactive vehicular applications. 99–106. https://doi.org/10.1145/1620509.1620529

[43] James A Russell. 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin* 115, 1 (1994), 102.

[44] Kashfia Sailunaz, Manmeet Dhaliwal, Jon Rokne, and Reda Alhajj. 2018. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining* 8, 1 (2018), 1–26. https://doi.org/10.1007/s13278-018-0505-2

[45] Karen L Schmidt and Jeffrey F Cohn. 2001. Dynamics of facial expression: Normative characteristics and individual differences. In *IEEE International Conference on Multimedia and Expo, 2001. ICME 2001.* Citeseer, 547–550. https://doi.org/10.1109/ICME.2001.1237778

[46] B. W. Schuller. 2008. Speaker, Noise, and Acoustic Space Adaptation for Emotion Recognition in the Automotive Environment. In *ITG Conference on Voice Communication [8. ITG-Fachtagung]*. 1–4.

[47] Amazon Web Services. [n.d.]. AWS Recognition API. https://docs.aws.amazon.com/rekognition/

[48] Mimi Sheller. 2004. Automotive Emotions: Feeling the Car. *Theory, Culture & Society* 21, 4-5 (2004), 221–242. https://doi.org/10.1177/0263276404046068

[49] Felix W Siebert, Michael Oehl, and HR Pfister. 2010. The measurement of grip-strength in automobiles: A new approach to detect driver's emotions. *Advances in Human Factors, Ergonomics, and Safety in Manufacturing and Service Industries* (2010), 775–783.

[50] A. X. A. Sim and B. Sitohang. 2014. OBD-II standard car engine diagnostic software development. In *2014 International Conference on Data and Software Engineering (ICODSE)*. 1–5. https://doi.org/10.1109/ICODSE.2014.7062704

[51] Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (2019), 50–55.

[52] Luke Stark and Jesse Hoey. 2021. The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 782–793.

[53] Ronnie Taib, Jeremy Tederry, and Benjamin Itzstein. 2014. Quantifying Driver Frustration to Improve Road Safety. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1777–1782. https://doi.org/10.1145/2559206.2581258

[54] Jianhua Tao and Tieniu Tan. 2005. Affective Computing: A Review. In *Affective Computing and Intelligent Interaction*, Jianhua Tao, Tieniu Tan, and Rosalind W. Picard (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 981–995. https://doi.org/10.1007/11573548_125

[55] ThoughtWorksArts. [n.d.]. EmoPy - Python Emotion Recognition Toolkit. https://github.com/thoughtworksarts/EmoPy

[56] Job Van Der Schalk, Skyler T Hawk, Agneta H Fischer, and Bertjan Doosje. 2011. Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion* 11, 4 (2011), 907. https://doi.org/10.1037/a0023853

[57] Marjolein D van der Zwaag, Joris H Janssen, Clifford Nass, Joyce HDM Westerink, Shrestha Chowdhury, and Dick de Waard. 2013. Using music to change mood while driving. *Ergonomics* 56, 10 (2013), 1504–1514. https://doi.org/10.1080/00140139.2013.825013

[58] Kiel von Lindenberg. 2014. Comparative analysis of gps data. *Undergraduate Journal of Mathematical Modeling: One+ Two* 5, 2 (2014), 1. https://doi.org/10.5038/2326-3652.5.2.1

[59] Heetae Yang, Jieun Yu, Hangjung Zo, and Munkee Choi. 2016. User acceptance of wearable devices: An extended perspective of perceived value. *Telematics and Informatics* 33, 2 (2016), 256–269. https://doi.org/10.1016/j.tele.2015.08.007

[60] Kangning Yang, Chaofan Wang, Zhanna Sarsenbayeva, Benjamin Tag, Tilman Dingler, Greg Wadley, and Jorge Goncalves. 2020. Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer* (2020), 1–20. https://doi.org/10.1007/s00371-020-01881-x

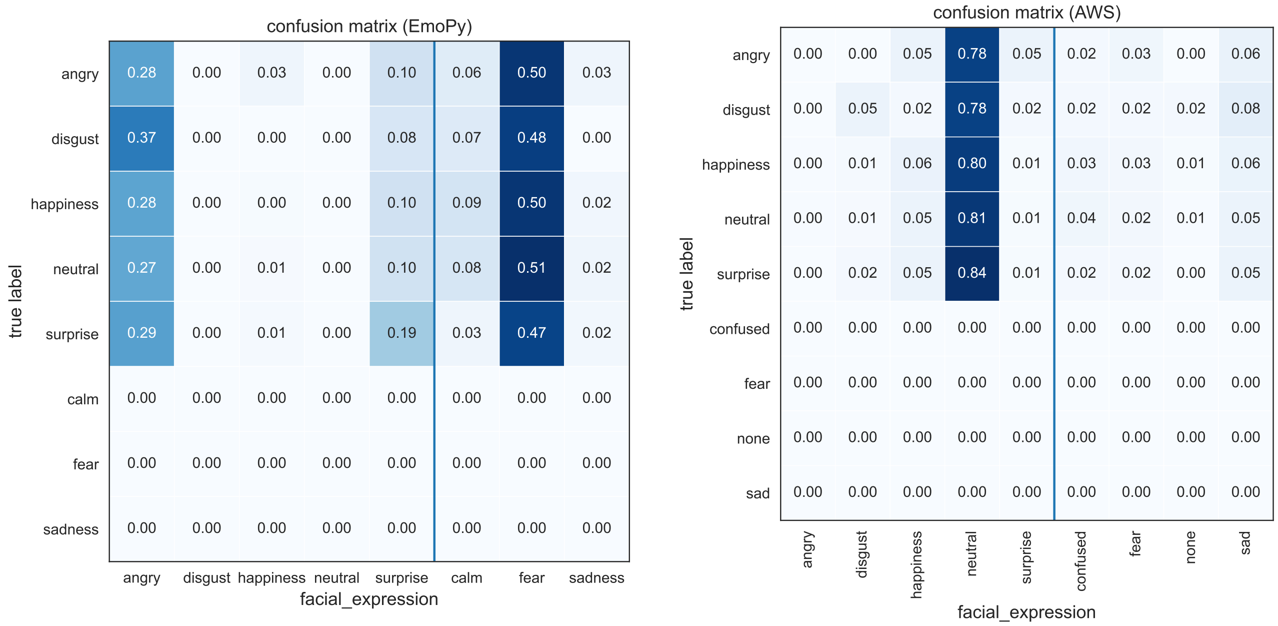
[61] Sebastian Zepf, Monique Dittrich, Javier Hernandez, and Alexander Schmitt. 2019. Towards empathetic Car interfaces: Emotional triggers while driving. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6. https://doi.org/10.1145/3290607.3312883

[62] Sebastian Zepf, Javier Hernandez, Alexander Schmitt, Wolfgang Minker, and Rosalind W Picard. 2020. Driver Emotion Recognition for Intelligent Vehicles: A Survey. *ACM Computing Surveys (CSUR)* 53, 3 (2020), 1–30. https://doi.org/10.1145/3388790

[63] Feng Zhou, Yangjian Ji, and Roger J. Jiao. 2014. Augmented Affective-Cognition for Usability Study of In-Vehicle System User Interface. *Journal of Computing and Information Science in Engineering* 14, 2 (02 2014). https://doi.org/10.1115/1.4026222 arXiv:https://asmedigitalcollection.asme.org/computingengineering/article-pdf/14/2/021001/6099446/jcise_014_02_021001.pdf 021001.

APPENDIX

A BASELINE FACIAL EXPRESSION ANALYSIS



(a)

(b)

Figure 6: Facial Expression analysis using a) publicly available facial expression analysis tool EmoPy b) cloud-service AWS Facial Recognition service. For AWS, we assigned 'calm' recognition labels to 'neutral'. Both classification system offer little predictive power in explaining perceived emotions on the ride. The accuracy overall of a) is 0.0076 **and b)** 0.5445**.**